DOCUMENT RESUME

ED 109 234 TM 004 695

AUTHOR Morrison, Max E.
TITLE Suggested Procedures for Developing an Efficient Testing Program.
PUB DATE Feb 75
NOTE 21p.

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Decision Making; Diagnostic Tests; Elementary Secondary Education; *Feedback; Instructional Improvement; Norm Referenced Tests; Standardized Tests; Student Evaluation; Student Needs; Testing; *Testing Problems; *Testing Programs; Test Interpretation; *Test Results; Tests; *Test Selection

ABSTRACT

Specifying the purpose and uses to be made of a test prior to administration is of paramount importance when one considers the time administrators and teachers spend on selecting, developing, administering and scoring tests and the amount of student time required to complete the tests. Determination of how test results will be used should be made prior to test development or selection. Possible questions needing answers may include: Is the purpose of the test to provide justification for a grade? Will the data be useful in diagnosing pupil strengths and weaknesses? How do the results from this test compare with those in other areas? Why do some pupils perform less well in one area than in others? Are these results acceptable or can something be done to improve future performance? Will the results aid in identifying pupils with special problems? Answers to the above questions may raise additional concerns for the test administrator and he may have to look to other information sources or devise other means to obtain relevant data on student performance. Before any formal testing gets underway, all available information on file should be reviewed to determine if additional data is necessary. To justify the test, it must be ascertained that more precise information is needed and that the data derived from the test will add something of significance to what is already known. (Author/BJG)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED109234

Suggested Procedures for Developing An Efficient Testing Program

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Max E. Morrison
Planning, Research, and Evaluation Division
Iowa Department of Public Instruction

TM 004 695

Introduction

The accountability movement has focused a great deal of attention to improved decision making for teachers and administrators through effective use of test results. Commercial test companies provide a number of options to test users. Such optional services include: item analysis, items keyed to objectives, listing of specific pages in a number of references for additional study if individual items or groups of items are answered incorrectly, and a number of other optional reports.

In addition to commercial test materials, teachers construct items to measure attainment of objectives they have set for their students. Most of the development efforts of teachers and commercial publishers have been centered on developing paper and pencil objective test items which can be readily scored. Primarily from this data, the teacher makes judgments about pupils. Decisions such as grouping, grading, remedial work, new assignments, and promotion and retention are typical examples. Test scores are recorded but little other data regarding the pupils' school experience are maintained. Such information as the ability to work with others, the amount of time spent on the learning task by the student, the initiative shown in completing a task, determination or perseverance to complete a task, identified areas of pupil interest, the ability to work independently and to communicate verbally may or may not be included in the data used to make judgments regarding individual students.

One problem prevalent with all machine scoreable paper and pencil tests is the time lapse between test administration and the reporting of results to the students. This practice continues even though research has consistently shown that immediate feedback directly affects the student's future test

performance. The student's desire to know answers is at a high peak following a test. Responding immediately to students' questions about test items missed may contribute to further learning and will help the teacher to identify questions most troublesome to the students. One of the shortcomings associated with standardized testing programs is the minimal feedback to the student regarding performance except in such general terms as a standard score or percentile rank with little or no discussion of the test items. This limited type of reporting often results in top performance from only highly motivated test takers.

Specifying the purposes and uses to be made of a test prior to administration is of paramount importance. When developing or selecting a test, it should be predetermined how the test results will be used. Possible questions needing answers may include: Is the purpose of the test to provide justification for a grade? Will the data be useful in diagnosing pupil strengths and weaknesses? How do the results from this test compare with those in other areas? Why do some pupils perform less well in one area than in others? Are these results acceptable or can something be done to improve future performance? Will the results aid in identifying pupils with special problems? Answers to the above questions may raise additional concerns for the test administrator and he may have to look to other information sources or devise other means to obtain relevant data on student performance.

This paper is like many tests in that it may raise more questions than it answers. When one considers the time administrators and teachers spend on selecting, developing, administering and scoring tests and the amount of student time required to complete the tests, then one must be concerned with what happens after testing. Administrators and teachers must carefully

review their own use of test results, continually seek better ways to use the data and/or discontinue some of the testing if the information is not being used. Having a folder full of test scores on each student contributes little to student learning unless meaningful decisions can be made to improve future educational experiences.

Before any formal testing gets underway, all available information on file should be reviewed to determine if additional data is necessary. To justify the test, it must be ascertained that more precise information is needed and that the data derived from the test will add something of significance to what is already known.

Why Test?

Teachers and administrators are constantly confronted with the need for accurate information on the performance of a student or a group of students. Information may be required for parent conferences, to identify students for special help or to inform the student of his progress. Information requirements of the teacher may be at one level while those of the administrator may be at another level. For example, the teacher may need to check on student progress in order to assign the next learning activity, while the administrator may require data from a number of classrooms to evaluate the appropriateness of curriculum materials.

Whatever the need for information, it should be clearly specified prior to any test administration if it is to be useful for purposes other than grading. Data collection should be limited to that which can be analyzed and interpreted by the teacher and/or the administrator. If one test administration can serve the needs of both users, it not only will be more cost effective but will allow additional time for instructional activities.

Questions for Administrators and Teachers to Consider Before Adopting a School Testing Program

To make effective use of test data, each decision-maker must specify the information needed prior to test selection and administration. Decision-makers may include parents, students, teachers, administrators, counselors and school boards. At least one unique piece of information should be singled out for each decision group and it should be specified how the data collected for one group relates to the total data collection.

Some general questions that need answers before any data collection plan is implemented include:

Is information needed on individuals or groups of students?

Do parents want to know how well their son/daughter is doing compared to students in other schools?

What are the perceptions of parents, teachers and students as to what the school is doing?

What are the expected outcomes by the teacher before the data is collected? Individual vs group.

Norm-Referenced Testing Programs

Specific questions on what data to collect and how to make effective use of the data for evaluation purposes include:

1) Is there an open discussion with faculty and/or students about the needs and aims of a measurement program?

2) How ready and willing is the staff to administer tests and interpret scores?

3) Are the objectives of the testing program clearly stated?

4) How can testing most effectively and efficiently contribute to the design of improved educational experiences?

5) Will testing after an educational experience focus on the information needs required for planning the next educational experience?

6) Will test results be studied longitudinally with information gathered systematically for teacher use and fitted into a well designed program?

7) Is the teacher familiar with the test instrument to be used?

8) Is the teacher involved in the selection of the test?

9) Is the teacher aware of the specific measurement characteristics of the test?

10) What are the limitations of the test?

11) Has the teacher read the publisher's statement of what the test is designed to measure?

12) Has the teacher read a description and evaluation of the test in Buros Mental Measurement Yearbook?

13) Has the teacher examined a specimen set of the items to be administered?

14) Has the teacher reviewed curricular objectives and subject content in the area(s) to be tested?

15) Has the teacher taken the test to check content validity?

16) What is the nature of the pupil population to be tested?

17) Will the test provide the student with information about himself in establishing realistic educational and vocational goals?

18) Will the test data be useful to parents in helping them understand that realistic goals must be set in the education of their children?

19) Will test data be used as a basis for discussion in inservice meetings?

20) What is the purpose of the test?

Placement - Selective grouping according to knowledge or skill in some subject area.

Diagnosis - To identify causes rather than outcomes.

Assessment - To collect census-like data on students to measure attainment of objectives related to knowledge, skills, and understandings.

Prediction - Predict chances for success in each course selected in the following year.

Evaluation - Study performance trends of pupils in certain subject areas.

21) What item analysis is planned following the testing administration to possibly reorganize the course of study?

22) What additional information will be needed to substantiate the test data?

23) What type of presentation is planned so that test results will be reported in language clearly understandable by students and parents?

24) What steps have been taken to insure test data will not be misused?

25) Have any plans been developed to identify teaching strategies for individuals, small groups, or the entire class for those objectives where students do not perform up to expectations?

26) Are students encouraged to question teachers and counselors regarding their performance on a particular test?

27) Are students encouraged to express their feelings in writing about test scores and the implications for future actions on their behalf?

Norm-referenced test data may be useful for administrators in making decisions regarding the following questions:

1) Does the data indicate any trends in student achievement?

2) How does local student performance compare with statewide performance?

3) Do pupils demonstrate greater competency at one grade level than at another?

4) Have competencies increased or decreased over the past three years?

5) Have previously identified weaknesses been resolved?

6) Do the competency levels vary from building to building?

7) Who and where are the pupils with special problems?

8) What skill areas need additional human and financial resources?

9) Should additional time be allocated to the basic skills in an effort to improve performance?

10) Should the expectation level for student achievement by the teaching staff and the community be revised?

Assessing a Local Testing Program

1) Who in the school system will use data from the standardized testing program? Teachers, students, counselors, administrators, parents and school board.

2) What is the primary purpose of each test administration?

3) What follow-up action was taken after the test administration last year? Item analysis, curriculum modifications, placement of students or similar type actions.

4) Is the testing program evaluated each year?

5) What processes are used to prepare students for a test?

6) What provisions are made to coordinate testing efforts across grade levels in the school?

7) How much teacher time (estimated) is spent in administering standardized tests? Include time spent on scoring, recording, and interpreting results to students.

8) What other information will be put together with the test scores prior to making a judgment?

9) Are all students tested or is a sample population of students tested to make inferences?

10) List all standardized tests (all types) administered last year).

Classroom Tests

The objectives of the test and those of the instructional program should be closely related. The teacher or administrator should identify crucial decisions where additional data is required and then select or develop a test to collect data to improve practices in the classroom or school. There are few available guidelines to direct the decision maker in this process.

Careful thought should be given when the major purpose of testing serves only to justify a grade. If this is the primary purpose, then standards of performance should be expressed for a particular grade and these should be consistent from teacher to teacher and from school to school. This would enable students interested in getting a good grade to concentrate on mastery of the skill or subject. Students no longer would have to "psych out" the teacher to determine what is essential to obtain a good grade.

As test results are seldom a true indicator of what a student knows or doesn't know, testing for grading purposes should be done frequently rather than at the end of a unit. Even then the results may not provide an accurate picture of a student's achievement, as some students may not be able to demonstrate mastery by means of a paper and pencil exam.

When the primary purpose of testing is to improve instruction, tests may provide information for a number of instructional and management decisions. Results from classroom tests may be used to place students in classes, identify students for additional testing and diagnosis, predicting chances for success in future courses or to assess student achievement before and after instruction.

Classroom testing should contribute to student learning through better planned learning activities designed to meet individual and group needs and

not be administered merely to assign a grade.

The following standards should be considered when developing, administering, and reporting test results:

1) Ease of test scoring should be only one of several items to be considered when selecting or developing a test. It is equally or more important to select or develop a test that can provide diagnostic data enabling the teacher to better assist the learner in mastery of the learning task.

2) The testing program should provide a number of ways for pupils to demonstrate mastery such as oral reports, individual tests, demonstrations, open book, etc.

3) The test should contain a balance of recall and application items as well as those requiring higher levels of thinking where the student must make a judgment.

4) Each test administration should include one open-ended response item, preferably a thought provoking question which will require a student to interpret, to apply and evaluate. The item should encourage the student to express some idea of his own.

5) The test should include items requiring transfer of skills from the classroom to real life experiences whenever possible.

6) Test items should be based upon the material covered by instruction and students should be prepared by being informed of the general purpose of the test. Purposes are to: inform parents of student progress, to diagnose strengths and weaknesses of the class, or to create student awareness of needs and progress.

7) Unanticipated outcomes from testing should be carefully observed. Are students frustrated, copying from other students, or making random responses?

8) An adequate reward system should be provided for all students. Constructive comments by the teacher on all tests may be adequate.

9) Immediate reporting of test results to students should be the rule of the day. The student is most interested in items answered correctly as well as items missed immediately following the test.

10) The possible benefits resulting from the test administration should always be weighed against the time required to carry out the testing.

11) The test data must be manageable for decision making. For example, an item analysis conducted following each test may yield more data than one teacher can work with effectively. The teacher may prefer to review data on a selected portion

of the class, i.e., those students who did not score as well as expected or the lowest 20 percent rather than attempt to analyze data from all students.

12) Observation of student behavior during test administration is essential to gain additional insight into students. This information may be helpful when one is analyzing a particular student's performance.

Possible instructional and management decisions that may result from an analysis of test data:

1) Assignment of additional drill and practice.

2) Review previous materials with individual or in small groups.

3) Review test items to reinforce concepts covered previously.

4) Assignment of library resources--magazines, papers and other books.

5) Instructor confers with individual(s) or small groups with similar problems.

6) Parent-teacher-student conference.

7) Use of programmed instruction for skill development.

8) Further diagnostic testing to identify causes for poor performance.

9) Recommend student(s) for remedial help.

10) Use other students to provide individual help or assign students to small group for additional help.

11) Refer student(s) to counselor, psychologist.

12) Retest--by using a sample of items from the original test, oral test or demonstration of proficiency in another way. Open-book test with another student.

13) Teacher may elect to spend additional time with student(s) experiencing problems.

14) Individual and/or small groups provided filmstrips, cassette tapes, games covering material presented.

15) Offer different reward system for part or all of the students.

16) Advance students to the next unit.

Preparing Test Items

The teacher should prepare test items prior to introduction of a new unit or develop them as the unit progresses to insure that the items are valid for the group taking the test. This process will insure that one is testing what has been taught rather than relying on memory of what should have been taught. Test items should be formulated from concepts which were thoroughly covered in class. If an arbitrary number of items must be answered correctly before a student may progress to the next instructional unit, this requirement should be clearly specified prior to the test administration.

Preparing Students for Testing

In preparing students for testing, the teacher should explain the general purpose to be served by the test. For example, if the test is to be administered to provide information for parent conferences, students should be so informed. If the purpose is to assist the student in becoming aware of his progress, then the teacher must follow-up the test administration through written comments or interviews with students not making the desired progress.

Test Administration

Feedback concerning successes and failures is an essential element in motivation. If maintaining a high degree of motivation for test taking is a goal of the school, then answer keys should be made available to students immediately after the test. It has been demonstrated that personal comments by the teacher regarding what the student did well and where improvement is desired tends to improve future performance more

than the mere assignment of a letter grade. Student motivation is enhanced when he has immediate feedback concerning his progress.

Purposes of Testing

Though the overall purpose of testing is to improve instruction, five intermediate purposes to achieve this goal can be identified. These purposes are not always discrete. For example, a decision maker may use diagnostic test results to place low scoring individuals for small group instruction within a class. Using diagnostic test information may eliminate the need for placement tests for this group of students.

Each of the five intermediate purposes for testing are described below.

Placement Tests

Placement tests provide information for assigning students to groups, classes, or courses. The test provides a gross level achievement score which is used to help identify the appropriate instructional level for the student. Other factors which are considered in placement include: student age, past performance and teacher judgment. Individual pupils are classified in relation to each other and in relation to the subject matter content.

Assessment

The main purpose of assessment is to compare pupil performance before and after instruction to determine the amount of growth or change taking place. Assessment tests may also be used to assess the effectiveness of teaching methods or instructional materials. It is concerned primarily with what has been learned.

Prediction

In prediction, the teacher estimates the capacity of students to achieve success in a course elected for the following year. The teacher makes the judgment based on the student's current performance. For example, it might be predicted that a good achievement test score in general science would lead to later success in physics. This use of test results should be carefully checked against empirical evidence to verify the correctness of the assumption.

Evaluation

Test results may be used to study trends or to report to the community on how well the students are performing. This judgment usually includes comparing a school's achievement test scores with achievement scores of a comparable group of schools.

Diagnosis

Classroom tests should be administered on a regular basis to locate weak spots in the new work that has been presented so corrective action can be taken at once. Day-by-day observations of the teacher may also provide insights regarding the students' problem so that action can be taken before it becomes serious. One objective of testing should be to help motivate students causing them to exert the necessary effort to master the task in the alloted time.

Specific diagnostic tests may be administered to determine where a weakness or deficiency exists. Diagnostic tests can be useful in pinpointing problems as they include a number of items for measuring each specific objective. A diagnostic test is designed to probe into the causes for poor

performance. The question needing answered must be specific, such as: "Does Johnny have the necessary vocabulary to read?", not "Why can't Johnny read?"

When a teacher analyzes diagnostic test results, he must consider the student's rate of work, his motivation, the accuracy of his responses and the difficulty level where most of the problems occur. Other evidence which may be contributing to the poor performance should be reexamined. For example, students who work rapidly and make numerous mistakes may do so due to carelessness, indifference or because of a lack of basic knowledge or skills. Further diagnosis may be in order, either through additional testing or via a teacher-pupil conference. Students who work slowly but are quite accurate may improve their performance through well-directed practice. The teacher may assign two or three students to a group to meet regularly to review the content covered in the test. Tutorial help either from the teacher, other teachers, aides, or student assistants should be provided whenever possible.

A number of approaches may be required to help students overcome learning difficulties. Improving work-study skills may be all that is required in some instances, while in others emotional and motivational problems may have to be dealt with in addition to work-study skills. Hopefully, diagnostic testing will enable the teacher to focus his teaching where it will produce the best results.

Diagnostic tests should be carefully chosen according to the specific information desired. Tests will identify the typical errors a student makes but the teacher must analyze the test results to identify possible causes for the errors. The test serves only as a starting point and additional supplementary information must be collected before an effective program can be designed for most students with learning difficulties.

To determine causes for persistent learning difficulties, the teacher may have to look not only at scholastic aptitude, mastery of the basic

skills and work-study skills, but at the home environment and the physical and emotional condition of the student.

In addition to analyzing diagnostic test results, diagnosis of learning difficulties should include information on a number of the following questions:

1) Approximately how much student time was spent on the learning task?

2) What types of assistance were provided to help the student with his problem?

3) Who provided the assistance?

4) What skill deficiencies must be overcome?

   Reading -- vocabulary
   comprehension
   word attack skills

   Listening

   Writing

5) Were there any unusual problems observed during the test?

6) Which students asked an unusually large number of questions?

7) Which students had difficulty concentrating on the test?

8) What probe questions were used to identify the source of the error?

   a) Did you understand the question?

   b) Why did you give this answer?

   c) Why did you leave the question unanswered?

   d) What gave you trouble on the question?

   e) How did you study for the test?

9) Was there any pattern evident in the questions missed?

10) Does the student have any physical or emotional problem?

Summary Comments

Whether testing requires five minutes or an hour, the cost of the test, the analyses of results, and the teacher's time required to administer the test must be considered relative to the use that can be made of the information.

If one of the major purposes of testing is to improve instruction, then some changes in test administration are long overdue. Administering a group test (norm-referenced or teacher made) to a room full of students is a very impersonal experience as all participants are treated alike. Many schools profess to individualize instruction but the test administrations completely ignore individual differences. The fact that some students work more slowly than others or that some are unable to work at a task for more than 15 minutes without some kind of break are not taken into account. Oftentimes the only feedback from this impersonal experience is a number (a raw score or a percentile score) and this may not come through for several days or weeks. Little or no personal interaction takes place between student and teacher with regard to the contents of the test. Frequently much of the information collected in the test is already available. Research has indicated that formal teacher judgment usually is in close agreement with the data gathered from a formal test. The value of collecting additional test information should be judged by how much it improves decisions over the best possible decisions made without it. In terms of cost effectiveness, it might be more profitable for the teacher to gather extensive data on those students needing special help rather than collecting data on all students. One must always keep in mind the possible unanticipated outcomes that may result from administering the test. (Does the student become less and less motivated to take tests? Does the student experience one failure after another? Does the student have inadequate time to complete the test?) These are but a few of

questions that should be asked by the test administrator, and answers can be obtained only through careful observation during the test and in analyzing and reporting of the results.

Prior to any test, the possible decisions that will be made from the data must be identified. In addition, the teachers and administrators must commit the necessary time and resources to do something with the results.

Those responsible for making decisions regarding students must take care to avoid becoming overly dependent upon paper and pencil testing. Other important learning outcomes (speaking and writing skills and work habits) should be observed and the information incorporated with test data before any judgments are made about an individual.

BIBLIOGRAPHY

American Psychological Association, Inc. Standards for Educational and Psychological Tests. Washington, D. C., 1974.

Anastasia, Anne. Testing Problems in Perspectives. American Council on Education. Washington, D. C., 1966.

Bloom, Benjamin S.; Hastings, J. Thomas; Madaus, George F. Handbook on Formative and Summative Evaluation of Student Learning. McGraw-Hill, Inc., 1971.

Boyd, Joseph L., Jr.; Shimberg, Benjamin. "Developing Performance Tests for Classroom Evaluation." Princeton, New Jersey, June 1971. (ED 052 259)

Cleary, T. Anne; Mathews, Walter M. "The Madison Plan: A New Approach to System Wide Testing." Symposium presented at annual meeting of the Wisconsin Educational Research Association, Milwaukee, 1971. (ED 064 335)

Clifford, Margaret. "Decision Making Rationale for Educational Testing." Iowa City, Iowa, October 1971. (ED 057 088)

Diederich, Paul B. "Short-Cut Statistics for Teacher Made Tests." Educational Testing Service. Princeton, New Jersey, 1973.

Educational Records Bureau. "Testing in Turmoil: A Conference on Problems and Issues in Educational Measurement." Greenwich, Conn., October 1970. (ED 054 202)

Educational Testing Service. "Making the Classroom Test: A Guide for Teachers." Princeton, New Jersey, 1973.

Educational Testing Service. "Multiple Choice Questions." Princeton, New Jersey, 1973.

Gronlund, Norman E. Measurement and Evaluation in Teaching. New York, N. Y.: The Macmillan Company, 1967.

Hoepfner, Ralph. "A Test of Tests." Center for the Study of Evaluation. Los Angeles, Calif., May 1971. (ED 055 112)

Kansas State Department. "A School District Testing Program Guide." Topeka, Kansas, 1971. (ED 066 442)

Katz, Martin. "Selecting an Achievement Test: Principles and Procedures." Educational Testing Service. Princeton, New Jersey, 1973.

Mauch, James E., and others. "The Gathering and Use of Data." October 1970. (ED 051 262)

Merwin, Jack C. "Testing for Testing's Sake." Invited address presented at the Annual Meeting of the National Council on Measurement, Chicago, Illinois, April 16, 1974.

National Education Association. "Guidelines and Cautions for Considering Criterion-Referenced Testing." A Working Draft of the NEA Task Force on Testing, June, 1974.

Venezky, Richard L. "Testing in Reading: Assessment and Instructional Decision Making." National Council of Teachers of English. Urbana, Ill., May 1974. (ED 091 669)

Webb, Eugene, et al. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago, Ill.: Rand McNally & Co., 1966.

Weber, George. "Uses and Abuses of Standardized Testing in the Schools." Occasional Paper Number 22, May 1974.

Weiss, David J. "Test Reviewing: Problems and Prospects." Inter-Association Council on Test Reviewing. (ED 044 439)